

DYNAMIC RESOURCE MANAGEMENT FOR DISTRIBUTED RETRIEVAL SYSTEM FOR SECURITY

FIELD OF THE INVENTION

[0001] The present invention relates to distributed architecture networks, and more particularly to resource management of a distributed architecture networked security system.

BACKGROUND OF THE INVENTION

[0002] In a distributed security system, various devices may be distributed on multiple networks. The components may be remotely located from each other or from a centralized controller. For example, cameras, monitors, recorder servers, meta-data servers, and archive servers may be attached to different networks. Often, the remotely-located components communicate wirelessly. For example, a network may use IP-based (Internet protocol) distributed architecture. The controller coordinates operations of the remotely located components. The security system activates recording or live analysis operations when a security event, such as motion, is detected. Because security events are not predetermined, conflicts may arise between various devices due to load and availability limitations.

[0003] It is therefore desirable to distribute the load on the devices according to various operations. The present invention uses a centralized controller to distribute operations to the devices according to load, availability, capacity, priority, and device capabilities.

SUMMARY OF THE INVENTION

[0004] A resource manager for a security system comprises one or more devices for collecting and/or managing data from an environment. One or more users submit operation requests for the data. A controller receives the requests and determines load characteristics of the devices. The controller allocates the devices to the operation requests according to the load characteristics.

[0005] In another embodiment of the invention, a resource manager for a security system network comprises a camera that collects multimedia data, a multimedia recorder that stores the multimedia data, and an analyzer that extracts meta-data from the multimedia data. One or more users submit operation requests for the data. A controller receives the operation requests and communicates with the camera, the multimedia recorder, and the analyzer to determine load characteristics. The controller allocates the operation requests according to the load characteristics.

[0006] In another embodiment of the invention, a method for allocating resources in a security system network comprises collecting data from an environment at one or more network resources. Requests are submitted for the data. Load characteristics of the network resources are determined at a controller. The network resources are allocated to the requests according to the load characteristics.

[0007] Further areas of applicability of the present invention will become apparent from the detailed description provided hereinafter. It should be understood that the detailed description and specific examples, while indicating the preferred embodiment of the invention, are intended for purposes of illustration only and are not intended to limit the scope of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The present invention will become more fully understood from the detailed description and the accompanying drawings, wherein:

[0009] Figure 1 is a functional block diagram of a distributed security system architecture according to the present invention;

[0010] Figure 2 is a functional block diagram of a resource manager according to the present invention; and

[0011] Figure 3 is a flow diagram of a resource scheduling and allocation service according to the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0012] The following description of the preferred embodiment(s) is merely exemplary in nature and is in no way intended to limit the invention, its application, or uses.

[0013] A distributed security network 10 includes an IP network controller 12 as shown in Figure 1. The IP network controller 12 communicates with one or more devices, including cameras 14, monitors 16, encoders 18,

decoders 20, a multimedia recorder 22, an analyzer 24, an IP gateway 26, an IP camera 27, and a PC controller 28.

[0014] The cameras 14 collect image data from an environment. The encoders 18 connect the cameras 14 to the IP network controller 12. The monitors 16 display the image data obtained from the cameras 14 in real-time. The decoders 20 connect the monitors 16 to the IP network controller 12. The multimedia recorder 22 records and archives the image data collected by the cameras 14. The multimedia recorder 22 plays the archived image data on the monitors 16 upon receiving a playback request. The playback operation is controlled by the user or analyzer unit 24.

[0015] The analyzer 24 analyzes the image data to extract meta-data. Meta-data includes the content, quality, and other characteristics of data. In the present invention, the meta-data may indicate security-related anomalies in the image data. One possible format of meta-data is MPEG-7. The analyzer 24 may analyze the image data in real-time or analyze recorded image data archived by the multimedia recorder 22. The analyzer 24 produces an alarm if an anomaly is detected in the image data. The PC controller 28 controls configuration characteristics of each device as well as the operations of each device.

[0016] Additionally, the security network 10 includes a meta-data server 30. The meta-data server 30 stores the meta-data associated with the archived image data. An initial meta-data entry includes, but is not limited to, recording purpose, date, time, and camera. The initial meta-data entry is created after the multimedia recorder 22 completes the recording of a multimedia stream

from the cameras 14. Additionally, the analyzer 24 may extract further information from the archived image data and update the meta-data entry in the meta data server 30. A user submits a query to retrieve meta-data from the meta-data server 30. The query returns location information for the requested archived image data. The location information includes a file descriptor and the corresponding multimedia recorder. The user then retrieves the archived image data from the multimedia recorder based on the location information. If the user accesses the security network 10 through the Internet, the user connects to the network 10 through the IP gateway 26. The IP gateway 26 connects the user to the proper source according to the operation. For example, if the operating requires retrieval of the real-time image data from a camera, the IP gateway 26 obtains the real-time image data of the camera and relays the media from the security network to the internet user's terminal.

[0017] A resource manager 32 allocates networked devices to the requested operations. The resource manager 32 manages the multimedia recorder units 22, the meta-data server units 30, and the analyzer units 24. Load on each unit is defined based on its functions in the security system. For example, the load of the multimedia recorder 22 is defined by available network bandwidth, available disk space, and available concurrency for the device. The number of concurrency denotes the number of concurrent operations on the device. For example, the multimedia recorder 22 may be able to handle a maximum of five simultaneous recording and playback operations. The load of the meta-data server unit 30 is defined by the available concurrency and

available disk space. The load of the analyzer server unit 24 is defined by the available bandwidth and available number of concurrency parameters. For each device, the PC controller 28 obtains the maximum value for each parameter during the device registration and informs the resource manager 32 about the availability and capabilities of each device. For example, the multimedia recorder 22 specifies its maximum network bandwidth, maximum concurrency, and maximum disk space information when the device is registered to the PC controller unit 28.

[0018] The resource manager 32 assigns networked resources to the operations requested by the PC controller unit 28. The resource manager 32 considers current resource availability, the resource requirement of the current operation request, and near future resource requests in determining how to allocate resources. The ability to consider near future requests allows the resource manager 32 to forecast the load on networks and devices in the near future. The PC controller 28 directs the devices that are selected by the resource manager 32 to perform operations.

[0019] The resource manager architecture 40 is shown in Figure 2. The resource manager 32 includes resource monitors 42 and a resource knowledge base 44. The resource knowledge base 44 includes a resource information service 46, a modeler service 48, a network proximity matrix 50, and a flow information service 52. The resource monitors 42 monitor the devices and network resources and update the resource knowledge base 44 with the information. For example, the resource monitors 42 may collect network

bandwidth usage information to reflect current traffic load on the network and update an entry associated with the monitored network in the modeler service 48. Devices provide their load information periodically. For example, a the multimedia recorder unit 22 provides the available bandwidth, average CPU utilization, and available disk space information periodically. In one embodiment, this information is included in a device keep-alive message, which is sent periodically, to reduce message traffic in the system.

[0020] Still referring to Figure 2, the resource manager 32 performs resource scheduling and allocation in response to a current request 54 according to near future requests 56 and current load on the managed devices. Requests may include a recording request, a live analysis request, an archive analysis request, a playback request, a play request, or a search request. The requests are triggered by alarms, triggered automatically according to a periodic schedule, or submitted by a system user.

[0021] The resource manager 32 allocates: multimedia recorder units 22 for recording or playback; meta-data server units 30 for searches; analyzer units 24 for live media analysis; analyzer units 24 and multimedia units 22 for archive data analysis; and multicast addresses for cameras 14. The resource manager 32 selects candidate devices based on the requested operation, the resource requirement of the requested operation, and the current load on the device and networks. In another embodiment, the size of a candidate set is limited so that only the first n devices are considered. The resource manager 32 evaluates each candidate device based on the requested operation, the resource

requirement of the requested operation, the current load on the device and networks, and the known future operation requests. The future operation requests are obtained from known periodic scheduled operations in the PC controller 28. During this evaluation phase, candidate devices will receive a score that defines order of preference between candidates. Then, the resource manager 32 communicates with the devices to reserve the operation on the selected device until one of the candidates accepts the operation, the candidate set is exhausted, or the operation times out. The resource manager 32 communicates with the devices sequentially based on an order defined by the scores. After a successful reservation, the resource manager 32 returns the selected device or devices as a response to the resource request. In another embodiment, the resource manager 32 directly sends the message to start the desired operation to prevent the reservation step. The resource manager keeps track of various load information associated with devices and networks that are described in the following sections.

[0022] The resource manager 32 builds the candidate devices set based on the requested operation type. In case of recording, playback, and live analysis operation, an element of the candidate set includes only one device. In case of archive analysis operation, an element of the candidate set includes two devices, namely the multimedia recorder 22, which will playback the archive media file, and the analyzer server 24, which will analyze the archived media file.

[0023] The resource information service 46 stores availability, current resource usage, maximum capacity, and current operations information for each

device. When the device is unregistered from the security system, the resource manager 32 moves the device to an unavailable state in the resource information service 46. When a resource allocation is completed, the resource manager 32 updates the entry associated with the device in the resource information service 46. For example, the resource manager 32 adds the information about the new operation, reduces the available bandwidth based on the operation, and updates concurrency values associated with the device. When an operation concludes on a device, the resource manager 32 updates the entry associated with the device in the resource information service 46 by removing the finished operation and updating the current resource usage information.

[0024] For each multimedia recorder 22, the resource information service 46 stores capacity information, including but not limited to maximum bandwidth, maximum disk space, maximum number of concurrent recording and playback operations, overwrite capabilities, and trans-coding capabilities. Overwrite capabilities are significant in the event that the hard disk is full. Trans-coding capabilities refer to the ability to convert between multiple media formats. The resource information service 46 stores the available bandwidth, available disk space, the available concurrency, average CPU utilization, and a record of each active operation in the multimedia recorder 22. The record of an active operation contains recording or playback operation details. The data related to recording operation contains a media source identifier, such as a camera identifier, the recording format, source media format, source media multicast address, source media bandwidth requirement, priority of operation, the

operation start time, and optional operation duration. The data related to playback operation contains the archive file identifier, playback media format, playback media bandwidth requirement, priority of operation, the operation start time, and receiving analyzer server, IP gateway, or user terminal information. When the multimedia recorder 22 has no disk space and the overwrite policy of multimedia recorder 22 is set to "not-to-overwrite", then the multimedia recorder 22 is not selected as a candidate device for recording operations. When the multimedia recorder 22 cannot have the proper trans-coding capability, then the multimedia recorder 22 is not selected for playback or recording even though the multimedia recorder has enough bandwidth and concurrency.

[0025] The resource information service 46 stores the following capacity information for each analyzer server 24: maximum bandwidth, maximum number of concurrent live and archive analysis operations, and available capabilities. The resource information service 46 stores the available bandwidth, the available concurrency, average CPU utilization, and record for each active operation in the analyzer 24. The record of active operation contains live or archive analysis operation details. The data related to live analysis operation contains media source identifier, source media format, source media multicast address, source media bandwidth requirement, priority of operation, the operation start time, and optional operation duration. The data related to archive analysis operation contains the archived file identifier, multimedia recorder 22 identifier, meta-data server 30 identifier, operation priority, and operation start time.

[0026] For each meta-data server, the resource information service 46 stores capacity information, including but not limited to maximum disk space and maximum number of search operations. The resource information service 46 stores the available disk space, the available number of concurrency, average CPU utilization, and record for each active search operation in the meta-data server 30. The data related to search operation contains priority of operation, and the operation start time.

[0027] The resource knowledge base 44 includes a modeler service 48 to build a graph representation of logical network topology for devices and networks in the security system. The graph contains two kinds of vertex definitions. A first type of vertex defines the sub-networks in the system. A second type of vertex represents the device on the network. The unidirectional edge between first type of vertices defines the connection from one sub-network to another. The bi-directional edge between second type of vertex and first type of vertex indicates that the device is connected to a sub-network. A device is added to the proper place in this representation when a device is registered and removed from this representation when a device is unregistered. The vertex representing the sub-network contains attributes such as subnet mask, gateway IP address, and maximum bandwidth. The maximum bandwidth is utilized in the resource selection phase to prevent the hot-spot on the network. This representation is used by network proximity matrix 50 and flow information service 52.

[0028] The resource knowledge base 44 includes a network proximity matrix 50 that allows the resource manager 32 to consider a cost of introducing a media flow which will go from one network to another network as a result of possible device allocation decisions. This cost value contributes to the score given to the candidate device if the device is not on the same network with the media source. Additionally, the cost contributes to the score if the network to which the candidate device is connected is not already receiving the media flow. For example, a monitor device 16 on the same network may receive this media flow. This matrix is dynamically configurable by the user or an application program such as a Network Manager System to influence allocation decisions made by the resource manager 32.

[0029] The flow information service 52 tracks information about active media flows in the security system. The information about an active media flow includes media flow statistics, such as format and bandwidth, source, and destinations of each active media flow. In the security system, the monitoring, recording, playback, live analysis, and archive analysis operations creates a media flow from a device to the plurality of devices. For example, the monitoring operation assigns cameras 14 to monitors 16 to display the camera image data based on a fixed or dynamic configuration. This creates a media flow from the camera site to the monitoring sites. When the periodic configuration is used, monitors 16 switch from camera to camera to scan the monitored environment. Thus, a plurality of monitors 16 may display the image data from the same camera 14. The recording operation introduces a media flow from camera 14 to

multimedia recorder device 22. During this live recording, another monitor 16 may be displaying the image data from the same camera 14. The live analysis operation introduces a media flow from camera 14 to analyzer server unit 24. The archive analysis operation introduces a media flow from the multimedia recorder 22 to the analyzer server unit 24. The playback operation introduces a media flow from the multimedia recorder unit 22 to the user's terminal or IP Gateway 26 when the user is accessing the system from the Internet. Information about the media flows is utilized to determine whether an allocation of device introduces an additional media flow in the security network. If the possible device allocation requires an injection of media flow into another network, the score of the candidate device contains a penalty based on the cost value obtained from the network proximity matrix 50.

[0030] The resource manager 32 provides a rule based configuration for system users for the selection of the multimedia recorder 22 and the meta-data server 30. These rules specify the user's preference about the recording place and meta-data storage place. The rule associated with recording place associates a set of multimedia recorders 22 with a specific camera 14 or camera group. The rule related to meta-data placement associates a set of meta-data servers 30 with a specific camera group. During the resource allocation phase, the resource manager 32 uses these rules to evaluate the candidate devices. For example, when the resource manager 32 needs to allocate a multimedia recorder 22 for a recording of camera 14 output, resource manager 32 considers the available resources on the devices as well as the rule associated with the

camera recording. Multimedia recorders 22 included in the rule that is associated with the requested camera 14 or camera's group receives a reward during the evaluation phase.

[0031] The RSAS 58 assigns a score to a candidate device according to the following formula:

$$S_i = W_{nc} * NCost(src, i) \\ + W_{dpl} * DataPlacementRule(src, i) \\ + W_{load} * (W_{bw} * BW_i + W_{ds} * DS_i + W_{conc} * Conc_i)$$

[0032] The W_{nc} , W_{load} , and W_{dpl} are weights that indicate importance of each parameter. The weights may be assigned or dynamically determined by user preference or network criteria. The src variable denotes the media source. The $NCost(src, i)$ denotes the value in the network proximity matrix 50. This value is normalized ([0..1]) and if the media flow associated with the source src is already in the network on which the candidate device i is connected, then it returns 0 to denote there is no cost. The $DataPlacementRule(src, i)$ returns 0 if the device i is not listed in the data placement rule associated with the camera src and 1 if the device i is included in the rule. The combination of BW_i , DS_i , and $Conc_i$ defines the load on the device i and each parameter is also associated with weight, W_{bw} , W_{ds} , and W_{conc} , respectively. The BW_i denotes the ratio of available bandwidth and maximum bandwidth on device i . The DS_i denotes the ratio of available disk space and maximum disk space on device i . The $Conc_i$ denotes the ratio of available concurrency and maximum concurrency of the device i . This weighted-sum formula produces a score value S_i for a candidate

device *i*. The RSAS 58 obtains the list of candidate devices from the resource information service 50 based on the current operation and desired requirement. For example, the live recording operation requires a multimedia recorder unit 22 while the live analysis operation requires an analyzer server unit 24. For example, the recording request of certain camera uses MPEG-2 while another uses MPEG-4 with different bandwidth requirement. If a multimedia recorder unit 22 has trans-coding capability, the selection process checks the multimedia recorder 22 capability against the required recording format in addition to the available resources of the candidate device. When there is a possibility of introducing a media flow into a network due to the candidate device, the resource manager 32 checks the maximum bandwidth, which is obtained from the modeler service 48, for the network to ensure that the network has enough bandwidth to handle the additional media flow. After the devices that are capable of handling the requests are selected, the RSAS 58 starts assigning the score value for each candidate device.

[0033] The resource allocation and scheduling service (RSAS) 58 performs a resource allocation and scheduling decision algorithm 60 as shown in Figure 3. A request queue receives and places the user requests based on the priority of submitted user requests at step 70. The RSAS 58 obtains near future requests and builds a set of requests at step 80. The set of requests includes a current request and the near future requests. The requests include information such as the desired operation, parameters of operation, and the required resources. Additionally, the requests may include the duration of the request.

The duration of a request is not included in user-controlled requests. The RSAS 58 builds a schedule for the set of requests by using the current resource information and network information from the resource information base 44 at step 90. For example, the RSAS 58 interacts with the resource information service 46 to determine which devices are capable of handling the desired operation. Additionally, the RSAS 58 interacts with the flow information service 52 to determine whether there is a media flow between a particular source and destination. The RSAS 58 interacts with the network proximity matrix 50 to determine the cost of introducing a media flow from the source to the candidate destination device.

[0034] The RSAS 58 uses the schedule to build a candidate resource allocation set at step 100. The RSAS 58 then evaluates the candidate set and ranks the candidates in the set based on scores obtained. After the set of candidate devices is evaluated, the RSAS 58 attempts to allocate the resources according to the rankings at step 120. At step 130, the RSAS 58 determines if the allocation was successful. If the targeted device or server is unable to accommodate the request, the RSAS 58 attempts to allocate the next candidate device from the candidate device set. If there are no more candidate devices in the candidate device set, then the RSAS returns a no resource error to the operation request at step 160. If the device or server accepts the request, the RSAS 58 commits the device or server to the requested operation. The RSAS 58 updates the resource information in the resource information base 44 and the

flow information service 48 at step 140 and restarts the algorithm at step 80 to process the next user request.

[0035] The description of the invention is merely exemplary in nature and, thus, variations that do not depart from the gist of the invention are intended to be within the scope of the invention. Such variations are not to be regarded as a departure from the spirit and scope of the invention.